

לימודי ערבית יהודית בעידן הבינה המלאכותית, סדנת מורים וחוקרים

ביום ה', 5 בפברואר 2026, נפגשנו ביד יצחק בן-צבי ליום עיון שעסק באפשרות לפתח לומדה ללימוד ערבית יהודית. השתתפו חוקרים, אנשי הוראה ומפתחים, והמפגש נועד בעיקר להיכרות עם המצב הקיים ולשיתוף ידע, ולא לקבלת החלטות. בפתיחה הוצגה המחויבות של מכון בן-צבי לשיתוף פעולה עם החברה לחקר התרבות הערבית-יהודית, וכן הוזכר העניין שמגלה מרכז צימבליסטה באוניברסיטת תל אביב בפרויקט. לאחר מכן עודד זינגר הציג את הרקע ליוזמה, שעלה כבר לפני כמה שנים מתוך ניסיון עם לומדות קיימות, וסקר חסמים שמונעים מסטודנטים להיכנס לתחום.

בהרצאה הראשונה משה לביא הציג תמונת מצב של התשתיות הדיגיטליות הקיימות. הוא עבר על מאגרי טקסטים (כמו פרויקט פרידברג, פרינסטון ופרויקט מדרש לתעתוק הגניזה), ועל כלים שונים שכבר נמצאים בפיתוח או בשימוש: מילונים דיגיטליים, כלי תעתוק (למשל עם מודלים לערבית יהודית), כלים למציאת מקבילות וציטוטים, המרת אותיות escriptorium מקלדות ייעודיות, ואפילו כיוונים לפיתוח מודלים של שפה. הוא הדגיש במיוחד את הצורך בזמינות דיגיטלית מלאה של חומרי היסוד, מילונים, דקדוק, טקסטים ומחקרים, כדי שהעבודה עם בינה מלאכותית תישען על תשתית אמינה ולא על חומרים אקראיים מהרשת. עלה גם רעיון לסמן למשתמשים את רמת האמינות של הטקסטים (טקסט מחקרי, טקסט אוטומטי, טקסט אוטומטי שעבר עריכה).

בהרצאה השנייה אבי ורשבסקי הציג את הניסיון של חברת mindcet בפיתוח כלים ללמידה דיגיטלית. הוא תיאר את המעבר ממכונות למידה קלאסיות, יקרות, לינאריות ונוקשות, לסוכני למידה מבוססי AI, שמאפשרים אינטראקציה פתוחה, התאמה ללומד ומגוון משימות. לצד היתרונות, הוא הצביע על בעיות: למידה כזו לא תמיד מייצרת שליטה עמוקה, והיא דורשת מהלומד לדעת לשאול את השאלות הנכונות ולהבין מה חסר לו. מרים גולדשטיין תיארה את הניסיון שלה בהוראה והדגישה שיש ביקוש ללימוד ערבית יהודית גם מחוץ לישראל (למשל באוקספורד וב-JTS). היא הצביעה על קושי מרכזי של סטודנטים: זיהוי רכיבים עבריים בתוך טקסט בערבית יהודית. לעיתים הם לא מזהים שמדובר במשהו שאינו ערבי, וגם כשהם מזהים—אין להם כלים להתמודד עם זה.

בדיון הפתוח ניסינו למפות את השאלות המרכזיות לקראת פיתוח אפשרי של לומדה. עלתה השאלה מי קהל היעד: חוקרים מהאקדמיה, חוקרי יהדות ללא רקע בערבית, דוברי ערבית ללא עברית, או קהלים נוספים. במקביל עלתה שאלה עקרונית אחרת, האם המטרה היא ללמד מיומנויות (קריאה, פענוח כתבי יד וכו'), או דווקא להנגיש תוכן וכלים שיבצעו את העבודה עבור המשתמש. בהמשך ניסינו לפרק את סוגי המיומנויות האפשריות: קריאה של טקסטים מודפסים, קריאה של כתבי יד, היכרות עם סוגות שונות, עבודה עם כלים דיגיטליים. כאן עלו שתי גישות: מצד אחד צורך במיקוד ברור כדי לא להתפזר, ומצד שני הצעה למודל רב-שכבתי שבו יש רמות שונות: הנגשה בסיסית, שכבת עיבוד ו-curation, שכבת למידה של מיומנויות ושכבת כלים מחקרניים.

בדיון עלו גם בעיות קונקרטיות: מחסור במילונים ובדקדוק לטקסטים מאוחרים, קורפוסים שלא זמינים ברשת, היעדר סטנדרט מוסכם לתעתוק, והפער בין שיטות תעתוק "נכונות" לבין מה שנוח לעבודה דיגיטלית. עלתה גם שאלת המוטיבציה ללימוד התחום במצב הנוכחי, וכן הצורך להסביר את ההקשרים הרחבים יותר של ערבית יהודית בתוך עולם השפות היהודיות. בסוף היום לא התקבלו החלטות, אבל כן התקבלה תמונה רחבה יותר של מה שכבר קיים ומה חסר. יצאנו עם הבנה שיש הרבה יוזמות וכלים, אבל הם מפוזרים ולא מחוברים, ושכדי להתקדם צריך גם מיפוי מסודר וגם החלטות לגבי כיוון, קהל יעד וסוג התוצר שרוצים לפתח.

Judeo-Arabic Studies in the Age of Artificial Intelligence: Workshop Summary

On Thursday, February 5, 2026, we convened at the Ben-Zvi Institute for a workshop devoted to exploring the possibility of developing a digital learning platform for Judeo-Arabic. The meeting brought together researchers, educators, and developers, and was intended primarily as an opportunity to survey the current state of the field and share knowledge, rather than to reach concrete decisions. The opening session presented the Ben-Zvi Institute's commitment to collaboration with the Society for the Study of Judeo-Arabic Culture, and also noted the interest expressed by the Zimblista Center at Tel Aviv University in the project. Subsequently, Oded Zinger outlined the

background to the initiative, which had already emerged several years earlier experience with existing digital learning tools, and reviewed barriers that hinder students from entering the field.

In the first lecture, Moshe Lavi presented an overview of existing digital infrastructures. He surveyed text repositories (such as the Friedberg Genizah Project, Princeton collections, and the Midrash Project for Genizah transcription), as well as various tools already under development or in use: digital dictionaries, transcription tools (e.g., eScriptorium with models for Judeo-Arabic), tools for identifying parallels and citations, letter-conversion utilities, dedicated keyboards, and even initial directions for developing language models. He emphasized in particular the need for full digital accessibility of foundational materials—dictionaries, grammars, texts, and scholarship—so that work with artificial intelligence can rely on a solid and reliable infrastructure rather than on random materials drawn from the web. The idea was also raised of indicating to users the level of reliability of texts (e.g., scholarly text, automatically generated text, automatically generated text that has undergone editing).

In the second lecture, Avi Varshavsky presented MindCET's experience in developing digital learning tools. He described the transition from classical learning systems—expensive, linear, and rigid—to AI-based learning agents that enable open interaction, personalization, and a wide range of tasks. Alongside these advantages, he pointed to certain challenges: such modes of learning do not always produce deep mastery, and they require learners to know how to formulate appropriate questions and to recognize gaps in their own understanding. Miriam Goldstein described her teaching experience and emphasized that there is demand for the study of Judeo-Arabic beyond Israel (for example, at Oxford and at JTS). She identified a central difficulty faced by students: recognizing Hebrew elements embedded within Judeo-Arabic texts. Often students do not realize that a given element is non-Arabic, and even when they do, they lack the tools to address it.

In the open discussion, we sought to map the central questions that would guide the possible development of a learning platform. One key issue concerned the target audience: academic researchers, scholars of Jewish studies without a background in

Arabic, Arabic speakers without knowledge of Hebrew, or additional audiences. A related conceptual question was whether the goal should be to teach skills (reading, deciphering manuscripts, etc.) or rather to provide access to content and tools that would perform these tasks for the user. We then attempted to break down the types of relevant skills: reading printed texts, reading manuscripts, familiarity with different genres, and the use of digital tools. Two contrasting approaches emerged: on the one hand, the need for a clear focus in order to avoid dispersion; on the other, a proposal for a multi-layered model comprising different levels—a basic access layer, a processing and curation layer, a skills-learning layer, and a layer of research tools.

The discussion also raised a number of concrete challenges: a shortage of dictionaries and grammars for later texts, corpora that are not available online, the absence of an agreed standard for transcription, and the gap between “correct” transcription methods and what is practical for digital work. Questions were also raised regarding motivation to study the field in its current state, as well as the need to situate Judeo-Arabic within the broader context of Jewish languages. By the end of the day, no formal decisions had been reached; however, a broader picture had emerged of what already exists and what remains lacking. We concluded with the understanding that, although many initiatives and tools are already in place, they are fragmented and insufficiently interconnected. Progress will require both systematic mapping and clear decisions regarding direction, target audience, and the type of product to be developed.